Journal of Computer Science 5 (11): 883-889, 2009
ISSN 1549-3636

Indoor Navigation and Localization for Visually impaired people Using Weighted Topological Map

Md Jan Nordin and Abbas M. Ali
Department of Computer Science, University Kebangsaan Malaysia,
Bangi, Selangor, Malaysia

Abstract: Problem Statement: Image base methods are a new approach for solving problems of navigations for visually impaired people. **Approach:** The study introduced a new approach of an electronic cane for blind people using the environment represented as a weighted topological graph instead, each node contains images taken at some poses in the work space, instead of building a metric (3D) model of the environment. **Results:** By computing weights between already stored images and the real scene of the environment and take some considerations like sessions. The system gives advices for the blind person to select the right direction in indoor navigate depending on weights and session, where a mono camera cane-held gives information in front of the visually impaired person. **Conclusion:** A cane that has the ability of getting SIFT feature for an object or site from a sequence of live images using the suggested approach is very satisfactory, The session and weight, speed up the system and gives a wide range indoor navigation and may be used to outdoor. Experimental results demonstrated a good performance of proposed method, the identification of different scenes to the blind person done by constructing the weighted visual environment graph to the system. The proposed scheme is using SIFT features to represent the objects and the sites.

Key words: SIFT, weighted topological graph, object recognition, sessions, localization

INTRODUCTION

Visually impaired people have one goal that to navigate through unfamiliar spaces without the human guide helps. To establish this navigation, they used many methods and devices, such as the long white cane and dog guide, to aid in mobility and to increase safe and independent travel as in[4,5,13,14]. There are two essential points to navigate the blind in the environment; first is adequate information about the travel path, so the blind person walking in a confidence and safe; and the second, is recognition of objects through the travel path which follow. These points get a wide area of interesting from the researchers. Towards that end, we proposed a model that can help visually impaired people to navigate any where indoor and outdoor sometimes, depending on the weight between the captured and stored scenes with in the session of the vision, where the new session of vision will be started with the relevant objects inside this session and the site will be identified to the mobile computer and also the blind person will be informed this session with this idea we have kitchen, bathroom, bedroom sessions. This increased independence, safety and confidence needed from the blind person.

To effectively navigate from one place to another, the blind person needs an internal representation or model of its environment.Sessions are important to describe scenes and matching or retrieving the stored scene much faster to convert it to the blind person as a voice scene. The most important point, that we use the proposed system in real time application for blind person. It will be miserable for the blind user, if the used system delayed while retrieving the stored object in the database. For this purpose, the use of sessions in the navigation gives advantage for the system, of dividing the database into parts according to the session. This will support the idea, that it is no need to find the car inside the kitchen room. This also gives intelligent behavior to the system and it will increase the speed of the system to remain in the real time in other side the blind person will be happy of the fast query answer of the system.

Related works: One of the main problems in computer vision and image processing is to perfectly recognize the object form its background. The blind person needs to recognize the object invariantly with occlusion,

Corresponding Author: Abbas M. Ali, Department of Computer Science, University of Kebangsaan Malaysia, Bangi, Selangor, Malaysia

different views (or scenes), scales, orientations and perspective transformations.

The basic method used for object recognition is to encode some of the properties of the object as a descriptor with it poses that can be stored in a database, this is used for localization process in the environment for the blind person.

The indoor navigation is heavily researched and the investigation continuous in this field, most of the research concentrate on the features of the scenes, the features are extracted in many ways, the common ways are Harries and SIFT extraction of image features, we shall concentrate in our study on SIFT where many navigations used SIFT feature descriptors as in [1,12,15], many papers proposed approaches to how to deal with the features extracted to be useful in the next stage of the system.

It is important to our system recognize the scene correctly, to do this we took what we see is distinctive and discriminative, so we don't incorrectly recognize scenes that are common in our environment. We established this by clustering descriptors; descriptors in a large cluster are less distinctive than those in a small cluster[7]. The Bag of Words algorithm has been applied to SIFT descriptors to identify discriminative combinations of descriptors [2,8,9]. For example when navigating indoors, window corners are common so are not good features to identify scenes with. Features found on posters or signs are much better, although even these may be repeated elsewhere. In[3] demonstrates real-time loop detection using a hand-held mono camera, using SIFT features and histograms (of intensity and hue) combined using a Bag of Words approach.

In[6] Also demonstrated real-time loop closure outdoors using SIFT features and laser scan profiles. Much work to remove 'visually ambiguous' scenes was needed and more complex profiles were preferred to provide more discriminative features.

The epipolar constraint does define an invariant feature, but this is defined by seven feature matches (up to a small number of possibilities) so eight or more points are needed for this to validate or invalidate a possible correspondence[16,10,2].

If suitable descriptors can be found then geometric constraints would be very useful for identifying distinctive scenes in an environment made up of different arrangements of similar components.

MATERIALS AND METHODS

The Smart Cane Perception (SCP) model: The cane in this work uses appearance based approach, by memorizing a set of images taken from the environment for the objects and their poses and entitled a weight according to its important for navigation.

A Cane-held mono camera, used for sending information in front of the blind person to the mobile computer when he/she need give order to analyze the view information, it is designed to involve the following stages in Fig. 1:

- Input stage, which is designed to import the objects along with its containing image view
- Calculate the local features for the image, containing the object by using SIFT algorithm and then calculate the weight for the scene
- Recognition Stage, through which, a matching decision is made when the input features are found close to an already existing image features in the database and according to the weight the system apply the localization process and will advice the blind which direction is a good direction to take. This could be considered as localization in the topological map of the environment. where includes the local landmarks after they detected in every image as described[7,11]

Localizing the blind person in such a topological map implies finding the node in the graph that resembles the current position depending on the weight.

A-the site and SIFT features: Since we are not controlling the blind person for any direction or focusing the cane in a specific view point. In SCP model the scenes treated as a set of features of SIFT and these features are having weights changing due to introducing new scenes of object appearance in the view side near from main view. So if we are taking more possibility of view side and calculating SIFT features including the objects and their poses for each scene and store it in the database, this will overcome

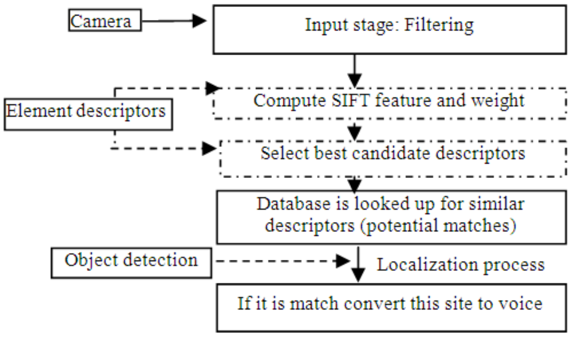
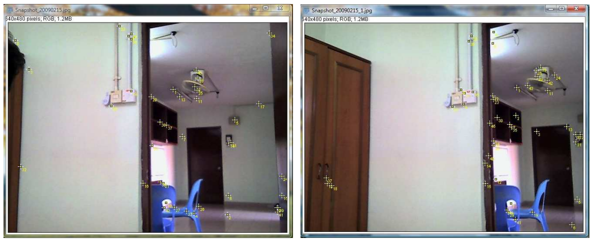
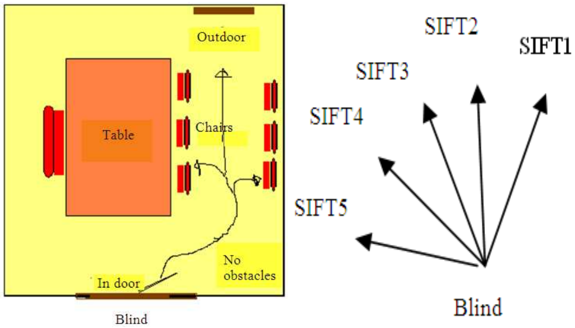

Camera
Input stage: Filtering
Element descriptors
Compute SIFT feature and weight
Select best candidate descriptors
Database is looked up for similar descriptors (potential matches)
Object detection
Localization process
If it is match convert this site to voice

Fig. 1: The general layout of the SCP model



(a) (b)

Fig. 2: The same site with different side view (a) is the same of (b) instead of counter part



(a) (b)

Fig. 3: The probability of SIFT sites for the blind person

the problem of the placing the cane from the blind person (position). Figure 2 shows two possibilities of the same site which can be near of the weights. And can give a good advice to the blind person to change the direction due introducing apart of counter in the second image.

In Fig. 2a and b. SIFT features extracted with Estimated transformation model: we have used 3×3 (Affine Transform). Some of these features extracted from the counter part. This however it is near to match Fig. 2a.

But these extra features it will decrease the competence of matching relative to other views. So here we tried to store the images which are near to others for the purpose of the weight of the similarity.

The search optimization for the site matching is very important to speed up the system SCP since SIFT algorithm is close to real time. So weights for each scene to give track information for the blind person it will optimize the calculation of the SIFT for features extraction. The SIFT algorithm is a four steps[1]:

- Scale-space extrema detection: Search over multiple scales and image locations
- Keypoint localization: Fit a model to determine location and scale. Select keypoints based on a measure of stability
- Orientation assignment: Compute best orientation(s) for each keypoint region
- Keypoint description: Use local image gradients at selected scale and rotation
- to describe each keypoint region

If the scales calculations near from the already stored image in the database for the same scale this will give the weight near form the scene so we shall give weight according to calculations of orientation and it shall take either less or more than the already stored features for the site (node in the graph); this will give the system more robustness. In Fig. 3 the blind person is in one of the node graph like office node and entering the office room, then a new session will be started with possibility of a number of sites directions as in Fig. 3b where S_i is SIFT (i) for the node n (i) means that the features for the site (i) for the system. So $S_i = \{fi_1, fi_2,\ldots\}$ and all sites in the known place is Ns so the probability of taking the blind person of the site (i) for node n (i) is:

$$P\ (Si/Ns) = P\ (Ns/S_i))*p\ (S_i))/P\ (Ns) \qquad (1)$$

where some of these sites may be not need from the blind person, it will be like the noise for the blind person.

It will be time consuming for the system, if the worst case happens. Where he time will be O (Ns)*O (S_i) this will lead the blind person to wait for the system answer however some times he/she can navigate without any information. To make the system friendlier with the blind person, the scalar weight from the image overlapped with image scene has been used, this let the system know which direction taken by the blind person and may be advice him/her of how rotate the stick to make it in the right direction. In general there are three ways to decrease the time complexity of the proposed system at all including the SIFT, to get faster answer for the blind person, these are:

Inside SIFT: This will happen by taking the idea that if the $S_i = \{Fi_1, Fi_2,\ldots\ldots,Fi_n\}$ and due to calculation of features Fi_1,Fi_2, if the number of features per octave (i) and octave (I + 1) near from any other octave (j) and octave (j + 1) in the database for any scene for node n (i) and weight $W_{n(i)}$, this will let us to take the remaining features calculations from the already stored image's octaves, this will decrease the time complexity for SIFT algorithm and it will be much suitable for the real time application like smart cane for the blind person. Where Lowe's algorithm depend on Gaussian function G for scaling space function L:

$$G(x,y,\sigma)=\frac{1}{\sqrt{2\pi\sigma}}e^{-\frac{1}{2}\frac{x^2+y^2}{\sigma^2}} \quad (2)$$

$$L(x,y,\sigma)=G(x,y,\sigma)*I(x,y) \quad (3)$$

and the DOG is a function give the difference between two scales σ and kσ:

$$\begin{aligned}D(x,y,\sigma)&=(G(x,y,k\sigma)-G(x,y,\sigma))*I(x,y)\\&=L(x,y,k\sigma)-L(x,y,\sigma)\end{aligned} \quad (4)$$

Scale space is separated into octaves; suppose that scaling take σ for octave1 and 2σ for octave 2 …. for the scene view within the node n (i):

$$G(x,y,\sigma,n(i))=\frac{1}{\sqrt{2\pi\sigma}}e^{-\frac{1}{2}\frac{(x^2+y^2)n(i)}{\sigma^2}} \quad (5)$$

$$L(x,y,\sigma,n(i))=G(x,y,\sigma n(i))*I(x,y) \quad (6)$$

Also the scale space it will be:

$$\begin{aligned}D(x,y,\sigma,n(i))&=G(x,y,\sigma,n(i))-G(x,y,\sigma,n(i))*I(x,y)\\&=L(x,y,k\sigma,n(i))-L(x,y,k\sigma,n(i))\end{aligned} \quad (7)$$

The key point's localization per octaves it will give us the object's image for node n (i). If its near from the stored key points localization for any object. Hence the system will not need make further calculation to get the best features for the image. Fig. 3b the SIFT features determined and stored in the database for each site of node n (i) with weight $W_{n(i)}$. when the blind entering the room the system directly help him/her to get the site with less calculation of the SIFT . Figure 4, shows less calculation of SIFT for some nodes.

Outside SIFT: This will happen by optimizing scenes for the SIFT algorithm to get the result as soon as possible this will be more time complexity because the time will be aggregated of (SIFT times)* number of applied images, till getting the best matching. It is obviously when we apply the weight, it will be decreased because the weight will give the indication of either the direction is left or right from the main view

Hybrid between inside and outside of SIFT methods: Our research is a study of how making the best matching of the objects in less time complexity to speed up the localization and help the blind person to navigate in a normal speed.

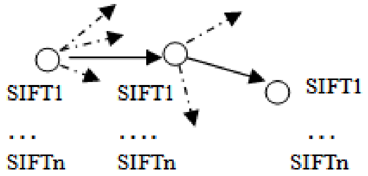


Fig. 4: Ready SIFTs for each node with weight directions

Calculating scalar weights: The similarity of two images is the degree of overlapped between them, if this degree is high, then a good matching between them will be done. the Scalar weight value for the matched image is expressed in Eq.8, where this value it has meaning whether it is positive, or negative, it will indicate concentrating of the SIFT features matched it left, center, or right.

$$W(fm(Sc_1, Sc_2)) = Sc_1(f_1,..\ f_n) \odot Sc_2(f1,..fn) \quad (8)$$

Function matching (fm) evaluated, if all f_i match for Sc_1 and Sc_2, then the scalar weight of the Sc_1 (real scene) it will be the same weight of the stored scene database (Sc_2) in this case the cane is the center of the scene. It is not need to advice the blind person to go to left or right. In the case that the real scene (Sc_1) not matched perfectly the scene database (Sc_2), in other words that some of features matches and some of them not match according to the threshold based in this system, then the blind person in the same session, but the Cane-held mono camera not towards the scene stored in the database so we should calculate the weight for this scene in the session. As we said, the weight depends on the features concentrated on which side of the scene. The considering of features concentrated in the left or right of the scene will be done according to threshold zone of the scene, then the scalar value depends on the probability of the scene Features. So the navigation of the blind person within the environment, depends on the weight and the important value for the scene with in the session. As an example if the scene it lead the blind person to go to another session this will be high value as in our example use it 8 otherwise we shall use to sets of weight values negative less than 8 it means go to left and positive more than 8 it means go to right. In general the weight value will give a good indication for the system to advise the blind person, for navigating or leaving the session, in the work space.

Experiment and Discussion: Suppose that each room of the environment is a node set of nodes (kitchen, bedroom, corridor) are V and suppose the way lead to these rooms are edges E and weight W, so the connected graph G = (V,E), dominate with the set of weights scenes for the room nc denotes the set of nodes of G, so suppose that ncj is a node so ncj = {{n1(0),wn1(0)}…..{n1 (i),wn1(i)}} and ncj+1 = {{n2(0),wn2(0)}….{n2(i),wn2(i)}}, it is clear that switching between ncj to ncj+1 it will be done in condition that the weight between them are high enough to convert the nodes. Then this it means that the images of the session will be change so the switching between nodes with weights gives wide range navigation to our system.

In this syudt we consider the general case when we start the session with a set of images taken at certain positions in the environment for this session and each link $e_{sn}ij$ in the graph denotes that scene i and j look similar in session e_{sn}. The similarity here is a scalar value of weight; we use to advice the blind person to perform navigation between two positions. and since we don't know which $e_{sn}ij$ selected to be navigated from the blind person we use the weight estimation of scenes this helped for localization and gave us support for a new approach of SLAM within the environment represented as sessions, so suppose that weight directions of the scenes at that node as in the Fig. 5 and the blind person direct the cane to another direction so the new scene weight will be counted and then inserted in the current session between the main directions.

Let us denote the 3D points in the current image for this session as $\{ei_1,ei_2,\ldots ei_n\}$ and the target as $\{ej_1,ej_2,\ldots ej_n\}$ and the weights $\{w(ei_1,ej_1), w(ei_2,ej_2),\ldots\}$,.as in the work of[5,10] we shall use essential matrix E to relate point correspondences:

$$(eim)^T E\, ejm = 0 \text{ for all m}$$

$$\text{where, } E = \begin{bmatrix} 0 & e2 & 0 \\ e4 & 0 & e6 \\ 0 & e8 & 0 \end{bmatrix}.$$

Blind Person navigation: the framework to navigate the blind person from one place (like the main gate) to another location (like the bed room), needs the images inside the database for that session in the environment mapped by the appearance based graph The general aim is that the blind person should be able to navigate to any room in a building by giving it a node in the graph according to the SCP advices. Where coming from room to another room is like going from one node to another in the graph case we desire that the system to change the active session to be ready to match the observation environment with set of session's images.

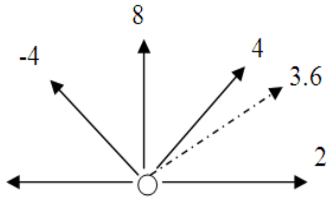


Fig. 5: The weight directions of nodes

One advantage of detecting the blind person current session is that it will inform the blind person the distances of images how long is far. So suppose that the bed room is 3×3 m^2 and the blind person entering the bed room so all the scenes inside the bed room is around 3m-ɕ where the ɕ is the position of the cane inside the room. Thus, the distance can be determined easily by using this method. There are many techniques exist to estimate the distance from two images. However these techniques would require us to make an assumption on the position of landmarks in the world, making the system less flexible.

One problem is that how the blind person take the shortest path to the goal location needed given by a node in the graph. First Dijkstra's shortest path algorithm[12] is used to compute the distance from every node n(i) in the graph to this goal node. This algorithm requires the links of the graph to be labelled with a distance measure while we have a similarity measure. The distances of the nodes to destination node are used during driving as a heuristic to drive in the direction of the selected node from the blind person.

Note, here we have many goals in our case and changes according to the blind person needed, so the algorithm in our case depend on table distance of the nodes, it can solve this problem. This it will also give the blind person some information and will avoid some path elements where the local features change rapidly (close to narrow throughways) and prefer to navigate in the centre of large open spaces.

The navigation procedure gives the advice to many nodes at a time that can be seen as sub goal nodes on a path to the goal node. This path of nodes could have been planned in advance. The Fig. 6 shows the path from main gate to bed room through corridor using weights.

However this would result in a very inflexible trajectory which would be difficult to traverse in a dynamic environment. The node of the graph with the highest similarity (The highest weight) is chosen as the current sub goal node of the blind person's cane. This procedure is linear in the number of nodes and could thus be time consuming. If a sub goal node is determined the SCP system tries to pick a new sub goal

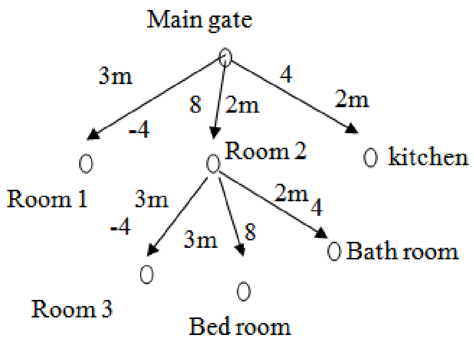


Fig. 6: Table distance points and weights

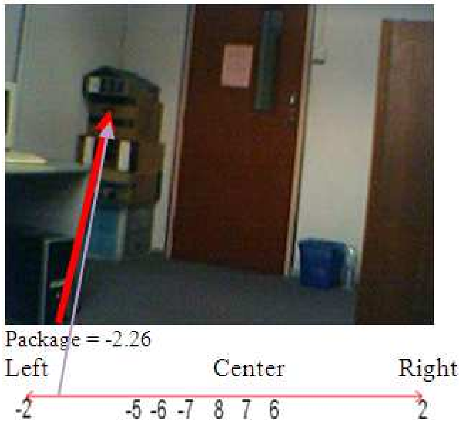


Fig. 7: Weights of the scenes match

by comparing the newest observation with all the neighbours. If one of these images matches, it becomes the new current sub goal c. until the blind person reach to the node wanted. Figure 7 shows the weights for two scenes.

RESULTS

In realistic scenarios for environment of the system, the blind person will navigate from one place (like the laboratory) through the corridor. The system matching stored object's images with the environment, inside the session, the scalar weighted matching scene supported the system for localization within the environment for the session. The system consistently shows good on-line performances for environments, the image for the object stored as features of SIFT to be directly match with the image scene used 320×240 jpg format, low quality with a good rate of correct recognition above 50% depending on the specific environment difficulties. The system meanwhile advise the blind person to navigate anywhere according to the internal process, with in the node in the graph. Figure7 shows the scalar weight given for the scene according to the SIFT feature store in the database.

CONCLUSION

A stick or cane that has the ability of getting SIFT feature for an object or site from a sequence of live images using the suggested approach is very satisfactory. Calculating SIFT features for random sites view and weights then storing them in the database helps to detect sites and objects. This speed up the algorithm and let the system to answering and help the blind person very quickly, as of real-time performance. Using the suggested approach, SCP model it will achieved quality of intelligent behavior for recognition the sites and objects since it will depend only on one camera and known value weights instead of extra calculations of the scenes octaves and scales is also higher compared to the other approaches.

The use of the Smart Cane Perception (SCP) Model suggested in our research differs from the other where this model helps to solve a problem associated with the object recognition. The SCP model not effected by illumination and scaling since mainly depend on the SIFT algorithm.

REFRENCES

1. David, G.L., 2004. Distinctive image features from scale-invarient key-points,cascade filtering approach. Int. J. Comput., 60: 91-110. http://direct.bl.uk/bld/PlaceOrder.do?UIN=150744203&ETOC=RN&from=searchengine
2. Booij, O., B. Tervijn, Z. Zivkovic and B. Krose, 2007. Navigation using an appearance based topological map. Proceeding of the IEEE International Conference on Robotics and Automation, Apr. 10-14, IEEE Xplore Press, Berlin Heidelberg New York, pp: 3927-3932. DOI: 10.1109/ROBOT.2007.364081
3. Filliat, D., 2007. A visual bag of words method for interactive qualitative localization and mapping. Proceeding of the IEEE International Conference on Robotics and Automation, Apr. 10-14, IEEE Xplore Press, Roma, pp: 3921-3926. DOI: 10.1109/ROBOT.2007.364080
4. Hub, A., J. Diepstraten and T. Ertl, 2004. Design and development of an indoor navigation and object identification system for the blind. Proceedings of the ACM SIGACCESS Conference on Computers and Accessibility, (CA'04), Atlanta, GA., USA., pp: 147-152. http://portal.acm.org/citation.cfm?id=1029014.1028657

5. Hub, A., T. Hartter and T. Ertl, 2006. Interactive tracking of movable objects for the blind on the basis of environment models and perception-oriented object recognition methods. Proceedings of the ACM SIGACCESS Conference on Associative Technology, Oct. 23-25, ACM Press, Portland, Oregon, USA., pp: 111-118. http://portal.acm.org/citation.cfm?id=1169007
6. Ho, K.L. and P. Newman, 2007. Detecting loop closure with scene sequences. Int. J. Comput. Vis., 74: 261-286.
7. Neira, J. and J.D. Tardos, 2001. Data association in Stochastic mapping using the joint compatibility test. IEEE Trans. Robot. Automat., 17: 890-897.
8. Fei-Fei, L. and P. Pietro, 2005. A Bayesian hierarchical model for learning natural scene categories. Proceedings of the IEEE Computer Society Conference on Computer Vision and Pattern Recognition, July 20-26, IEEE Computer Society, Washington DC., USA., pp: 524-531. http://portal.acm.org/citation.cfm?id=1069129
9. G. Csurka, C.R. Dance, L. Fan, J. Williamowski and C. Bray, 2004. Visual categorization with bags of keypoints. Proceedings of the IEEE Workshop on Statistical Learning in Computer Vision, (SLCV'04), pp: 1-16. http://www.cs.cmu.edu/~efros/courses/LBMV07/Papers/csurka-eccv-04.pdf
10. Booij, O., Z. Zivkovic and B. Krose, 2006. Pruning the image set for appearance based robot localization. Proceedings of the 12th Annual Conference of the Advanced School for Computing and Imaging, June 2006, The Netherlands.
11. Sonnenblick, Y., 1998. An Indoor Navigation System for Blind Individuals. Proceedings of the 13th Annual Conference on Technology and Persons with Disabilities, Feb. 28-28, pp: 147-152. http://www.dinf.ne.jp/doc/english/Us_Eu/conf/csun_98/csun98_008.html
12. Jauregi, E., J.M. Martinez-Otzeta, B. Sierra and E. Lazkano, 2007. Handle identification by keypoint extraction. Robotics and Autonomous Systems Group University of Basque Country, Donostia. http://www.sc.ehu.es/ccwrobot/publications/papers/jauregi07handle.pdf
13. Tissot, N., 2003. Indoor navigation for visually impaired people-the navigation layer. ETH Zürich Technical Report. http://www.mics.ch/SumIntU03/NTissot.pdf
14. Hub, A., J. Diepstraten and T. Ertl, 2004. Augmented indoor modeling for navigation support for the blind. University of Stuttgart, Universitätsstraße 38. http://www.pubzone.org/dblp/conf/cpsn/HubDE05
15. Mozos, O.M., A. Gil, M. Ballesta and O. Reinoso, 2008. Interset Point Detectors for Visual SLAM", Recent Advances in Control Systems, Robotics and Automation, 2nd Edn., International Sar. ISBN: 978-88-901928-3-8, pp: 190-199.
16. Booij, O., Z. Zivkovic and B. Krose, 2007. Image based navigation using a topological map. Proceedings of the 13th Annual Conference of the Advanced School for Computing and Imaging, June, 2007, The Netherlands.